DOCUMENT RESUME

ED 351 371 TM 019 195

AUTHOR Blankmeyer, Eric
TITLE L-Scaling: An Update.
PUB DATE [92]
NOTE 17p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; Equations (Mathematics); *Mathematical Models; *Multivariate Analysis; *Sampling; *Statistical Inference; *Weighted Scores
IDENTIFIERS Data Reduction Methods; Leontief Matrix; *L Scaling; Sensitivity Analysis; *Weighting (Statistical)

ABSTRACT

L-scaling is introduced as a technique for determining the weights in weighted averages or scaled scores for T joint observations on K variables. The technique is so named because of its formal resemblance to the Leontief matrix of mathematical economics. L-scaling is compared to several widely-used procedures for data reduction, and the discussion proceeds in terms of descriptive statistics since the various techniques have sampling properties that are either unknown or intractable. In principle, a researcher should choose a scaling method by proposing a model that explains how the discrepancies arise. This inferential approach is difficult in cases where the X data do not satisfy such requirements as multivariate normality and the statistical independence of the observations. In these cases, a researcher may choose to apply a kind of sensitivity analysis by comparing the outcomes of several scaling methods, including L-scaling. Four tables present analysis results. (SLD)

# L-scaling: an Update

Eric Blankmeyer
Department of Finance and Economics
Southwest Texas State University
San Marcos, TX 78666

1. Introduction.

Given T joint observations on K variables, it is frequently useful to consider the weighted average or scaled score:

$$y_t = \Sigma_k X_{tk} w_k \ , \ t = 1,\ldots,T \ .$$

In matrix notation,

$$y = Xw \ = XWe \quad . \qquad (1)$$

In expression (1),

X = a TxK data matrix to be scaled (the input);

y = a column vector of T scaled scores (the output);

w = a column vector of K weights;

e = a column vector of K units (1's); and

W = a KxK diagonal matrix whose nonzero elements are the weights (w = We).

This paper introduces L-scaling as a technique for determining the weights. The technique is so called because of its formal resemblance to the Leontief matrix of mathematical economics. L-scaling is compared to several widely-used procedures for data reduction, but no attempt is made to survey the voluminous literature on scaling methods. The discussion proceeds in terms of descriptive statistics since the various techniques have sampling properties that are either unknown or intractable.

To deal with the "apples and oranges" problem that arises in scaling incommensurable variables, it is assumed that the data have been standardized. That is,

$$R = X'X \qquad (2)$$

is a correlation matrix of order K. (This premise is relaxed in section 5, where equivariance is discussed.) Another assumption

is that the K variables are not perfectly correlated: the rank of R exceeds 1. In applications, the rank of R is usually the smaller of T and K since there is unlikely to be an exact linear relationship among the variables.

2. L-scaling.

Because the variables are imperfectly correlated, there are potentially TxK discrepancies between the weighted average y and its components XW. In view of expression (1), L-scaling defines such a discrepancy as $X_{tk}w_k - y_t/K$. In matrix notation, the TxK discrepancy matrix is

$$\begin{aligned} D &= XW - ye'/K \\ &= XW - XWee'/K \quad \text{from (1)} \\ &= XW(I - ee'/K) \quad , \qquad (3) \end{aligned}$$

where I is the identity matrix of order K. L-scaling chooses the weights to minimize the sum of the squared discrepancies. In other words, the weights minimize the trace (tr) of D'D, just the sum of that matrix's diagonal elements:

$$\begin{aligned} tr(D'D) &= tr\{[XW(I - ee'/K)]'[XW(I - ee'/K)]\} \\ &= tr\{XW(I - ee'/K)][XW(I - ee'/K)]'\} \quad (4) \end{aligned}$$

since in general tr(PQ) = tr(QP) for conformable matrices. Moreover, (I - ee'/K) is an idempotent matrix, so expression (4) becomes

$$tr(D'D) = tr[XW(I - ee'/K)WX'] \quad . \qquad (5)$$

In expression (5), the t-th diagonal element of the bracketed matrix is

$$\Sigma X_{tk}^2 w_k^2 - (1/K)\Sigma\Sigma X_{tj}X_{tk}w_j w_k \quad , \quad (6)$$

where the summations over j and k run from 1 to K. Since the X data are standardized, it follows from expression (6) that the

L-scaling minimand is

$$tr(D'D) = w'(I - R/K)w \quad , \qquad (7)$$

where R is defined in expression (2) and w = We is the column vector of K weights.

To avoid the trivial solution (w = 0), expression (7) must be minimized subject to a normalization of the weights. L-scaling adopts the constraint that the weights should add to 1:

$$w'e = 1 \quad . \qquad (8)$$

Whether the constrained minimum is unique depends on the rank of (I - R/K) = (KI - R)/K. The matrix is evidently singular if and only if K is an eigenvalue of R. But then the rank of R is 1, contrary to assumption; and the K variables collapse to a single variable. Barring this, the rank of R exceeds 1, the inverse of (I - R/K) exists, and the L-scaling minimum is unique. This conclusion is valid whether or not $T \geq K$ and even if some (but not all) of the X variables are linearly dependent.

When the quadratic form (7) is minimized with respect to w and subject to the normalizing constraint (8), the L-scaling weights are

$$w = c(I - R/K)^{-1}e \quad . \qquad (9)$$

In expression (9), the positive constant

$$c = 1/e'(I - R/K)^{-1}e \qquad (10)$$

makes the weights add to 1. In addition, c is the value of the quadratic form (7) at its constrained minimum. Substitution of the weights into expression (1) produces the scaled scores y.

3. L-scaling and the Leontief matrix.

In many applications of scaling, all the correlations are positive; in other words, the K variables tend to rise and fall together. While L-scaling can certainly be applied in other situations, it will be assumed in this section that R is a positive matrix.

In that case, the array (I - R/K) bears a formal resemblance to the Leontief matrix that has a prominent role in the theory of linear economic models. Such matrices are positive definite. Moreover, they have positive elements on the principal diagonal and negative elements elsewhere. Hawkins and Simon (Ref. 1) show that these properties guarantee a strictly positive inverse:

$$(I - R/K)^{-1} > 0 \quad . \qquad (11)$$

It follows from expressions (9) and (10) that the L-scaling weights are also strictly positive. Blankmeyer (Ref. 2) gives a concise proof of the Hawkins-Simon result.

Waugh (Ref. 3) shows that the Leontief inverse can be expanded in power series. For L-scaling the expansion is, apart from the factor c,

$$y = X(I - R/K)^{-1}e = Xe + XRe/K + XR^2e/K^2 + \ldots + XR^ne/K^n + \ldots \quad , \qquad (12)$$

where n is a positive integer. The sequence converges since $Re/K < e$.

The first term in the sequence is Xe, just the row totals of the data matrix. The n-th term in the sequence approximates the largest eigenvector of R if n is a large integer. Accordingly, the L-scaling solution subsumes two well-known scaling techniques: simple row means and the first principal component of

the correlation matrix. The relationships among these scaling methods are further developed in the next section.

4. L-scaling and other techniques.

Table 1 provides a direct comparison of three multivariate methods: L-scaling, the first principal component, and what Raj (Ref. 4, 16-17) has called the best weight function. (While each method generally leads to a different solution, the symbols w and y are used for all three methods to simplify notation.) Several comments may be helpful.

(1) In all three methods, the scaled scores are computed as y = Xw once the weights have been obtained.

(2) The L-scaling criterion was introduced in section 2. It provides a least-squares fit between a scaled score $y_t$ and each of its weighted components $X_{tk}w_k$; there are potentially TxK such discrepancies. Under principal components, a least-squares approximation to the X matrix is the matrix yw', whose rank is 1 and which gives a row-and-column representation of X. Again, there are TxK discrepancies. The best weight function minimizes the variance of the scaled scores (whose means are zero); this least-squares problem involves just T discrepancies.

(3) The choice of a normalization rule is important. If either L-scaling or the best weight function is minimized on the unit sphere (w'w = 1) rather than on the plane (w'e = 1), the principal-components solution is obtained. In particular, the weights that minimize on the unit sphere

$$\begin{aligned} & w'(I - R/K)w \\ = \; & w'w - w'Rw/K \\ = \; & 1 - w'Rw/K \end{aligned} \qquad (13)$$

evidently minimize -w'Rw or equivalently maximize w'Rw.

(4) Both L-scaling and principal components provide solutions as long as the rank of R exceeds 1. The best weight function, however, requires the inverse of R, which implies that the rank of $R = K \leq T$ . This is a limitation. For example, if 10 cities were to be ranked on the basis of 15 quality-of-life variables ($T = 10$, $K = 15$), the best weight method could not be used to obtain a scaled score for each city.

(5) If all correlations are positive, L-scaling and the first principal component have positive weights; but the best weight function may have zero or negative weights. In some applications, negative weights may make the results hard to interpret.

(6) As long as the scaling problem is subject only to a normalizing constraint, computer solutions for all three methods are straightforward. L-scaling and the best weight function require inversion of a KxK matrix, while the weights for the first principal-component are calculated by raising R to a sufficiently large power. In some applications, however, it may be useful to apply linear constraints (equations or inequalities). For example, one might want to know how all the scaled scores are affected when the third observation is ranked a priori at least as high as the seventh: $y_3 \geq y_7$ or equivalently $\Sigma(X_{3k}-X_{7k})w_k \geq 0$. Under such constraints, L-scaling and the best weight function become exercises in quadratic programming, for which algorithms are available. On the other hand, it might be less straightforward to compute the first principal component subject to a set of linear inequalities.

(7) When the data matrix X may be contaminated by outliers, a

robust scaling technique is required. An approach which retains all the algebraic properties of L-scaling is the weighted least-squares minimand (Ref. 5):

$$\Sigma\Sigma(X_{tk}w_k - y_t/K)^2 H_{tk} \quad , \qquad (14)$$

where the weight $H_{tk} = 1/|X_{tk}w_k - y_t/K|$ unless the discrepancy is zero, in which case $H_{tk} = 0$. Expression (14) is therefore equivalent to:

$$\Sigma\Sigma|X_{tk}w_k - y_t/K| \quad , \qquad (15)$$

subject to the T+1 constraints $Y = Xe$ and $w'e = 1$. As a multivariate version of a median, expression (15) is relatively resistant to outliers. The solution may be obtained by linear programming. If the dual form is applied and the upper-bound constraints are handled implicitly, the problem involves just TxK+1 non-negative variables and K explicit constraints [Wagner (Ref. 6)]. At the maximum of the dual linear program, the shadow price of constraint k is the weight $w_k$. The initial simplex tableau is described in Table 4.

(8) Perhaps the simplest scaling method of all is row means ($y = Xe/K$), where each weight is set equal to 1/K without regard to the information contained in the correlation matrix. When are equal weights optimal ? All three methods summarized in Table 1 produce equal weights if the correlations among the K variables happen to be identical. The methods of Table 1 also produce equal weights if the correlation matrix exhibits a pattern like the example in Table 2, due to Morrison (Ref. 7, 245-246). Unless R displays such regularities, at least approximately, the equal-weight solution may provide a poor fit in comparison with the other methods discussed in this section.

(9) If R has rank K, there are K distinct principal components. Together, they reproduce R, accounting for all the correlation among the variables. Can L-scaling also generate several indices from the same data ? Having once calculated y and w, one can compute the discrepancy matrix D in equation (3) and replace R by D'D in equations (9) and (10). This leads to a second y and w, and the steps can be repeated. Unlike principal components, the various L-scaling indices are not orthogonal and do not reproduce R. In this respect, L-scaling more nearly resembles the factor-analytic methods used in psychology and sociology, where allowance is made for sampling error. In factor analysis, one hopes to explain most of the correlation structure, but one does not expect to account for all of it in a mechanical way.

5. Equivariance.

Index numbers measure to what extent several variables move in lockstep. In other words, do all the variables tend to change in proportion ? It is reasonable to require that this proportionality be preserved after a rescaling of some variable(s). Geometrically, the plane of best fit still passes through the origin; the rescaling should merely alter its tilt. However, it is not reasonable to expect that proportional variation will survive an arbitrary shift in the zero of one or more variables, since the plane of best fit is then displaced from the origin, contrary to the hypothesis of proportionality.

Accordingly, the scaling techniques in Table 1 are not --and should not be-- equivariant for shifts in the origins of the data. The measurement of proportional variation logically

requires a decision about the appropriate zero of each variable.

As to the effect of rescaling a variable, the first principal component is altered drastically:

"This dependence on the unit of measurement is obviously a weakness of the principal component technique....If a variable is measured in such small units that its numerical values dominate those of the other...variables, the first principal component will reflect the value of this variable rather closely...." [Theil (Ref. 8), 55].

Nor does the use of a correlation matrix really avoid the dilemma, for standardization is itself a choice of units. There are, after all, many ways to make the data dimensionless. (For example, one might divide each variable by its mean.) Each rescaling leads to a different principal-component index, and the various indices may give subtantiallly different impressions of the degree of proportional variation.

Of course, factor analysis <u>is</u> invariant to any single-valued transformation of the variables. However, the many proposals for "rotating" the factor-analytic solution show that there remains a fundamental indeterminacy about the choice of units.

Unlike the first principal component, the L-scaling index adjusts in a simple way when a variable undergoes a change of units. Let us abandon the assumption that the data have been standardized. It is still true that the L-scaling matrix is obtained when each diagonal element of X'X is multiplied by (1-1/K) and each off-diagonal element is multiplied by -1/K. The resulting matrix is then inverted; and the L-scaling weights are just the row sums of the inverse matrix, normalized to add to

one.

Now suppose that the first variable in X is rescaled. Each observation on the first variable is multiplied by some positive constant, z. This means that the first row of the L-scaling matrix is multiplied by z; next, its first column is multiplied by z. No other element of the matrix is changed. The end result is that the first row of the inverse matrix is multiplied by 1/z; next, its first column is multiplied by 1/z.

How do these operations affect the L-scaling index, y ? Since it is obtained by multiplying the inverse matrix into the unit vector, the index y is unchanged so long as the first element of the unit vector is replaced by z.

More generally, the unit vector is to be replaced by ($z_1$, $z_2$,..., $z_K$) when each of the K variables is rescaled. (Under this renormalization, the L-scaling weights no longer add to one.)

The situation for L-scaling and the best weight function may be summarized this way: indices computed before and after a change of units are identical if one adopts the renormalization outlined above. Of course, it would usually be pointless to change units and then undo the job by renormalizing. Rather, this discussion is intended to show that, in L-scaling, nothing essential is involved in the choice of units. The same cannot be said with respect to principal components.

6. A simulation and some conclusions.

As an hypothetical example, 100 observations on three variables were drawn from a pseudorandom-number generator (Ref. 9, seed = 8445). That is, T = 100 and K = 3. Specifically, the data matrix was computed as:

$$X(t,1) = G(t,1)$$

$$X(t,2) = G(t,1) + G(t,2)$$

and $$X(t,3) = 4G(t,1) + G(t,3)/G(t,4) \qquad (16)$$

where t = 1, ... , 100. The G's are independent standard normal variables. The first and second X variables are therefore normally distributed. However, the observations on the third X variable are expected to contain outliers since the ratio G(t,3)/G(t,4) is a Cauchy random number with an indefinitely large variance.

Based on the standardized values of the three X variables, Table 3 displays the empirical correlation matrix for the sample of 100 observations together with the weights for the three methods of Table 1 and for the robust version of L-scaling in equation (15). The four sets of weights differ notably from one another, and it follows that the indices would also differ. Under the robust version, the third X variable has a large weight because its outliers are ignored.

In principle, a researcher should choose a scaling method by proposing a model that explains how the discrepancies arise. However, this inferential approach is difficult in cases where the X data do not satisfy such requirements as multivariate normality and the statistical independence of the observations. In view of these obstacles, a researcher may choose instead to apply a kind of sensitivity analysis by comparing the outcomes of several scaling methods, including L-scaling which has been introduced in this paper.

Table 1. Comparison of 3 scaling techniques

| | Minimand | First-order condition | Normalization |
|---|---|---|---|
| L-scaling | $\Sigma\Sigma(X_{tk}w_k - y_t/K)^2 = w'(I - R/K)w$ | $(I - R/K)w = e$ | $w'e = 1$ |
| First principal component | $\Sigma\Sigma(X_{tk} - y_t w_k)^2 = -w'Rw$ | $(\mu I - R)w = 0$ | $w'w = 1$ |
| Best weight function | $\Sigma_t(y_t)^2 = \Sigma_t(\Sigma_k X_{tk}w_k)^2 = w'Rw$ | $Rw = e$ | $w'e = 1$ |

Note: $\mu$ is the largest eigenvalue of R.

Table 2. A patterned correlation matrix

| | | | |
|---|---|---|---|
| 1.00 | | | |
| 0.70 | 1.00 | | |
| 0.60 | 0.40 | 1.00 | |
| 0.40 | 0.60 | 0.70 | 1.00 |

Table 3. Weights for a correlation matrix

Correlation matrix

| | | |
|---|---|---|
| 1.000 | | |
| 0.726 | 1.000 | |
| 0.184 | 0.134 | 1.000 |

| Weights | L-scaling | First principal component | Best weight function | Robust L-scaling |
|---|---|---|---|---|
| W1 | 0.368 | 0.419 | 0.230 | 0.234 |
| W2 | 0.363 | 0.413 | 0.313 | 0.231 |
| W3 | 0.269 | 0.168 | 0.457 | 0.535 |

Note: the weights for the first principal component are renormalized from w'w=1 to w'e = 1 to facilitate comparison with the other three sets.

Table 4. Initial Simplex Tableau

The tableau may be characterized as follows:

- o Number of variables (all non-negative) = TK + 1.
- o Number of explicit constraints = K.
- o Right-hand side of each constraint is ≤ 0.
- o Maximize variable number TK + 1.
- o Upper bound of 2 on each variable except number TK + 1.
- o For constraint 1:

| Variable number | Left-hand side coefficient |
|---|---|
| 1 | $(K-1)X_{11}$ |
| 2 | $-X_{11}$ |
| ... | ... |
| K | $-X_{11}$ |
| K+1 | $(K-1)X_{21}$ |
| K+2 | $-X_{21}$ |
| ... | ... |
| 2K | $-X_{21}$ |
| ... | ... |
| TK+1 | 1 |

Table 4 (concluded)

o For constraint 2:

| Variable number | Left-hand side coefficient |
|---|---|
| 1 | $-X_{12}$ |
| 2 | $(K-1)X_{12}$ |
| ... | ... |
| K | $-X_{12}$ |
| K+1 | $-X_{22}$ |
| K+2 | $(K-1)X_{22}$ |
| 2K | $-X_{22}$ |
| ... | ... |
| TK+1 | 1 |

o For remaining K-2 constraints, pattern of coefficients analogous to constraints 1 and 2.

References

1. HAWKINS, D. and SIMON, H., Some Conditions of Macroeconomic Stability, Econometrica, Vol. 17, pp. 245-248, 1949.

2. BLANKMEYER, E., Approaches to Consistency Adjustment, Journal of Optimization Theory and Applications, Vol. 54, No. 3, 1987.

3. WAUGH, F., Inversion of the Leontief Matrix by Power Series, Econometrica, Vol. 18, pp. 142-154, 1950.

4. RAJ, D., Sampling Theory, McGraw-Hill, New York, New York, 1968.

5. SCHLOSSMACHER, E., An Iterative Technique for Absolute Deviations Curve Fitting, Journal of the American Statistical Association, Vol. 68, pp. 857-859, 1973.

6. WAGNER, H., Linear Programming Techniques for Regression Analysis, Journal of the American Statistical Association, Vol. 56, pp. 206-212, 1959.

7. MORRISON, D., Multivariate Statistical Methods, McGraw-Hill, New York, New York, 1967.

8. THEIL, H., Principles of Econometrics, Wiley, New York, 1971.

9. DOAN, T. and LITTERMAN, R., Regression Analysis of Time Series (RATS), VAR Econometrics, Minneapolis, Minnesota, 1984.